

VIEWPOINT

All Is Not Well in the World of Translational Research

Ellis F. Unger, MD

Silver Spring, Maryland

It is not unusual for novel treatment strategies to fail in clinical trials, despite highly encouraging results in pre-clinical proof-of-concept studies. Typically, such “failures of translation” are blamed on the poor predictiveness of animal models. Often, however, the poor predictiveness of today’s preclinical proof-of-concept studies is related not to limitations of the models but to investigator bias and a lack of scientific rigor. The resulting false-positive results only serve to mislead the field and impede medical progress. With the resurgence of translational research, it is useful to examine some of the problems that plague these studies and consider their solutions. With thoughtful planning, execution, and analysis, it is possible to generate reliable and predictive data from preclinical proof-of-concept studies, results that should more rapidly advance medical progress. (J Am Coll Cardiol 2007;50:738–40) © 2007 by the American College of Cardiology Foundation

Melvin Marcus was a giant in academic cardiology, a critical thinker who appreciated the nuances of preclinical studies. At scientific meetings, he would provide thoughtful analyses and critique to junior (and not-so-junior) scientists. Frequently, he would surprise a speaker (and audience) by identifying a crucial issue that had been overlooked by the investigators, throwing a study into doubt. He mentored countless young investigators with this insight and wisdom. Tragically, he succumbed to cancer in 1989, at the age of 49 years.

It is a different time now. Presentations of study results, including truly fantastic findings, go largely unchallenged at scientific meetings; in fact, they seem to be readily accepted as the norm. Investigators induce experimental myocardial infarction in rats, and subsequent injection of stem cells into the tail vein leads to a 30% improvement in left ventricular ejection fraction! Transfer of a gene encoding an angiogenic cytokine improves collateral perfusion by 40%. The pattern has been the same for many years now: acceptance of highly encouraging results without a thoughtful critique or careful weighing of study limitations. Similarly, outstanding pre-clinical results often are presented in publications with little in the way of critical analysis. Viewed optimistically, these “proof-of-concept” studies can serve as a launching pad for Phase 1 studies in human subjects. If the drug or biotechnology product appears to be reasonably well tolerated in Phase 1, possibly with some anecdotal evidence of biological activity, a Phase 2 clinical program is initiated. At some

point in late-Phase 2, when it becomes clear that the product has no beneficial effect in patients, we blame the animal model for deceiving us! We declare that the model was a poor predictor of human disease.

Much has been written regarding the limitations of animal models as predictors of responses in humans. Certainly, animal studies have unavoidable shortcomings, but many of their limitations originate, not from the models themselves, but instead from “human factors,” i.e., the design and interpretation of these studies—by humans.

What Is Wrong With Translational Research?

Whole-animal cardiovascular research was an area of intense interest in the 20th century, importantly advancing our understanding of cardiovascular physiology and pharmacology. But with exciting and rapid advances in molecular biology in the 1980s, whole animal studies seemed suddenly passé. Costs associated with large animal research skyrocketed. Research grants became more difficult to obtain. A heightened awareness of animal welfare posed additional challenges. Bright and promising cardiovascular fellows were encouraged to seek training in molecular biology, not animal physiology. Well-established investigators in animal research moved on to pursue other endeavors, and the pool of qualified mentors dwindled. Now, at least in part because of these factors, expertise in preclinical animal investigation seems sorely limited. Fewer individuals demonstrate the capability to perform adequate preclinical “proof-of-concept” studies. This lack of proficiency appears to go unrecognized, possibly because relatively few individuals grasp the issues well enough to provide substantive critique. Although it may be a bit extreme to declare that the “blind are leading the blind,” it does not seem far-fetched to suggest that the “myopic” are leading the “myopic.” Thus,

From the Division of Cardiovascular and Renal Products, Office of New Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland. The opinions expressed herein are those of the author and do not necessarily reflect those of the Food and Drug Administration.

Manuscript received April 5, 2007; revised manuscript received April 18, 2007, accepted April 24, 2007.

despite the enthusiasm for the resurrection of translational research, there is a paucity of individuals with the capacity to perform or interpret these studies competently.

What Can Be Done to Improve Preclinical Translational Research?

The preclinical “proof-of-concept” study is a critical element of translational research. Using an animal model designed to mimic a human disease, the demonstration that a drug or intervention has a salutary effect on a surrogate end point provides the necessary support to move the treatment strategy forward into the clinic. Preclinical studies can be largely exploratory, designed to generate hypotheses, in which case they can hardly be called “proof-of-concept,” but they may be nevertheless important. Conversely, well-conceived and designed preclinical investigations, carefully executed and rigorously analyzed, may be suitable for formal hypothesis testing. The latter type of study is more likely to accurately model human disease and less likely to misinform.

The principal challenges in designing, conducting, and analyzing preclinical translational research are familiar to clinical trialists: variability and bias. Fortunately, with only moderate effort, both can be reasonably managed in preclinical studies.

Minimizing variability. The tension between variability and effect size is common to all translational research: to detect the consequences of an intervention, the effect size must overcome the variability in the model. In contradistinction to clinical studies, which typically enroll hundreds or thousands of subjects, preclinical studies with as few as a dozen animals may be adequately powered to detect the effect of a treatment strategy. The relative homogeneity of the animals and lack of confounding factors underlie this important difference. In contrast to clinical studies, many aspects of preclinical studies can be carefully controlled, and it is advantageous to strive for consistency in methods and analyses whenever possible. Such consistency serves to limit variability.

Disease-free animals should be derived from a single source, and should be approximately the same age. (A case can be made for selecting animals of a single gender, depending on whether gender effects are expected.) All animals should be exposed to the same handling, housing, bedding, food, and water. If surgical procedures are required for the model, the same individual, or the same team of individuals, should conduct the surgery in all animals, particularly if the techniques are complex (e.g., survival surgery). If refinements in surgical technique or improvements in other procedures are anticipated, the operator (or team) should hone his or her skills before initiating the controlled study. A single individual, or a consistent team of individuals, should conduct all treatments and all end point assessments.

Bias: the obvious and the not so obvious. Investigators are under tremendous pressure to achieve “positive” results. Positive results have the potential to lead to publications that enhance professional recognition and bring about

personal advancement. Negative study results may (or may not) be extremely important from a scientific standpoint, but they do little to advance one’s career. Moreover, it is difficult to convince editorial boards of medical journals to accept negative studies for publication (negative publication bias). In the case of translational research, proving the feasibility of an original approach to the treatment of a disease provides yet another inducement to achieve positive results, and proof of feasibility may lead to patents and lucrative licensing arrangements. These incentives lead to powerful biases, both conscious and subconscious, with the potential to undermine a study. Fortunately, the application of some of the principles of design and analysis familiar to clinical trialists can largely neutralize these biases.

Randomization and blinding. It goes without saying that preclinical proof-of-concept studies should include a concurrent control group, almost without exception. Historical or nonconcurrent control groups are of limited value in establishing proof-of-concept. Randomization and blinding are critical to the interpretation of these studies. Variable block randomization is reasonable for small animal studies; investigators do not need to be appraised of the randomization scheme. Disparate study agents (active agents and vehicle) should be indistinguishable, if possible. An experienced investigator who is not directly involved in the project should prepare the randomization code and fill a set of coded, labeled vials. They should retain the password-protected treatment code in a secure location. (It is also advisable for this individual to mail the randomization code to an independent party, as a backup in case of unforeseen circumstances.) Each animal should receive its treatment assignment in sequence, matched to a specific coded vial(s).

It is critical to maintain rigorous blinding of all investigators throughout the study. Thus, selection of particular animals and application of surgical techniques (if applicable) should be carried out without knowledge of treatment assignment. It is imperative that: 1) investigators are blinded to the identities of study agents; 2) investigators are blinded when end points are assessed; and 3) investigators remain blinded during data processing and analysis. This author has heard investigators explain that their study end points are entirely objective and therefore not susceptible to bias. Although this may be true on occasion, there are additional, more subtle reasons, as discussed herein, for maintaining a blinded study.

Missing data. Typically, complex preclinical investigations suffer from missing data (e.g., animal deaths, technical issues). It is a fact, however, that small studies are exquisitely sensitive to the effects of missing data; the inclusion or exclusion of a single data point can alter the results and conclusions of a study. Thus, it is critical to have a prospective plan to deal with missing data and outliers. If decisions must be made regarding missing data, they must be made without knowledge of treatment assignment. Although it is less obvious, missing data also may arise by choice. For example, decisions regarding ancillary animal care and treatment are sometimes required during the course of a study, and they have the potential to

affect the study results. Consider a strategy that is hoped to improve left ventricular function after experimental myocardial infarction. Confronted with an animal in florid heart failure, treatment decisions have to be made for the benefit of the animal. Some might argue that it is best to euthanize the animal. Given an animal with a predictably poor study outcome, knowledge of treatment assignment could most certainly bias such decisions. However, even if the investigator is blinded, the judgment to euthanize the animal is tantamount to consciously changing an outcome from “poor” to “missing.” Depending on the plan for handling missing data, these decisions may bias the study and undermine the study results and conclusions.

Data outliers pose yet another problem. Consider once again the study of acute myocardial infarction. The recording of a left ventricular ejection fraction of 95% in an animal with a sizable myocardial infarction is probably spurious. An ejection fraction of 70% may be spurious. Is there some cutoff of unreasonableness, beyond which outliers should be eliminated? If so, standard methods to identify outliers should be established prospectively. Finally, in publications and presentations, animal deaths and the handling of missing data and outliers should be discussed with the study results, if applicable.

In summary, missing data and outliers can strongly influence the results of preclinical studies. If there are prospective plans in place to address these issues, and if decisions are made without knowledge of treatment assignment, there is less opportunity for bias to influence the study.

Statistics. It is not unusual to observe misuse of statistical tests in preclinical studies. For example, use of paired *t* tests to analyze unpaired data, and use of parametric methods to analyze non-normally distributed data are not uncommon. Multiplicity is often an issue in preclinical studies that is inadequately addressed. Consider a study with 2 active treatment groups (2 different routes of administration) and a vehicle control group, with 2 main end points, each assessed at 3 points in time. Under these circumstances, there are 12 ways to “win,” and a rigorous statistical approach requires adjustment of alpha. However, there is nothing inherently wrong with exploring all possibilities and presenting nominal *p* values, as long as the data are interpreted fairly under the guise of an exploratory study. Alternatively, the prespecification of a single primary end point and a single time for its assessment would eliminate much of this multiplicity, such that the statistical plan would only have to account for the fact that there are 2 active treatment groups.

Study limitations. All studies have limitations. The typical reader may not be an expert in the field and may lack the

sophistication necessary to ferret out a study’s pitfalls. The limitations section of the paper should present a clear, honest, informative, and balanced view of the data and their interpretation. A less-balanced “devil’s advocate” perspective on the study’s limitations and interpretations can be very informative as well. If animals died or were otherwise excluded, this result should be explained. Missing data that may have compromised the results should be discussed. If this was an exploratory study (e.g., if 5 end points were analyzed without priority and without adjustment of alpha) this should be stated. If there was potential unblinding during the course of the study, perhaps because of a blood pressure-lowering effect of the active treatment, this should be explained. The usual “disclaimer” language, that the particular animal model may have limitations in predicting effects in humans with disease, may be true, but it is not particularly useful.

Substantiation of findings. Positive findings in a single study should be independently substantiated by a subsequent study. Given the strength of the biases that are operational during preclinical proof-of-concept studies (both conscious and unconscious), the frequency of a false-positive result may be substantially greater than the 5% “guaranteed” by a nominal *p* value of 0.05. These biases exist despite the best intentions of the investigators. Substantiation is neither as impractical nor as difficult as it sounds. Subsequent studies can be designed to extend the results of a previous study, while also substantiating them. By examining alternative modes of drug delivery, altering the timing of drug delivery or, in particular, studying higher or lower doses of the study agent, critical incremental knowledge can be gained, while at the same time substantiating the initial results. There is also much value in comparing the test agent head-to-head with another agent that has known biological activity. Although such follow-up studies may seem inherently less exciting than the initial study, they are nevertheless important.

Conclusions

With careful planning, rigorous execution, and thoughtful analysis, valid and persuasive data can be obtained from preclinical proof-of-concept studies. Through the application of the aforementioned principles, false-positive results will be less likely to misguide investigators, and the field will be advanced more rapidly.

Reprint requests and correspondence: Dr. Ellis F. Unger, HFD 110, Building 22, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993. E-mail: ellis.unger@fda.hhs.gov.
